Business Analytics (110-1)

Assignment 2

1.

GoodBelly is a company doing business mainly in the United States. We may see GoodBelly's sales outcome in different regions, in different stores, in different time periods, and with different marketing activities from the dataset "`GoodBelly_data.csv`". The data came from 126 Whole Foods stores over the 10 weeks between May 4 and July 13. There was a total of 1,386 observations. The definition of the variables in the dataset are as follows.

> **1** *Weekly Sales (Volume):* The number of units sold per store per week.
>
> **2** *Average Retail Price:* The average retail price for GoodBelly products per store per week.
>
> **3** *Sales Rep:* Defined as 1 if the store had a regional sales rep (face-to-face contact) and 0 if the store had only the national sales rep (no face-to-face contact).
>
> **4** *Endcap:* Defined as 1 if a store participated in an endcap promotion.
>
> **5** *Demo:* Defined as 1 if the store had a demo on the corresponding week.
>
> **6** *Demo1-3:* Defined as 1 if the store had a demo 1-3 weeks ago.
>
> **7** *Demo4-5:* Defined as 1 if the store had a demo at least 4-5 weeks ago.
>
> **8** *Natural Retailers:* The number of other natural retailers within 5 miles of each store.
>
> **9** *Fitness Centers*: The number of fitness centers within 5 miles of each store.

We would like to see whether those marketing efforts are effective and worthwhile. Investigate this data set and use the methods introduced in class to address the following questions.

(a) Do descriptive statistics, with R, to provide an overview for your retailing business. You may do it from any perspective with any EDA method. You may include some basic summaries as well as some emphases on interesting findings.

(b) Based on your findings from (a), comment on the marketing activities about their effectiveness. Use some graphs and numbers to support your comments. You may comment on all of them, rank them, making suggestions about how to use them. Of course, your comments may be different from region to region, from time to time, or depending on any factor that you find useful.

(c) Build a linear model to explain the relationship between sales and promotional efforts, and interpret the regression output.

(d) Does the in-store demo program boost the sales? If so, for how long does the sales lift last?

(e) Does the placement of the product within the store affect the sales?

(f) What other factors affect the sales of Goodbelly's products? Based on the regression output, what are your recommendations to Goodbelly's management?

(g) Are there any suggestions to improve and refine the model?